International Journal for Parasitology 41 (2011) 1185–1195

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Testing Mendelian inheritance from field-collected parasites: Revealing duplicated loci enables correct inference of reproductive mode and mating system

Jillian T. Detwiler *, Charles D. Criscione

Department of Biology, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:
Received 10 May 2011
Received in revised form 12 July 2011
Accepted 14 July 2011
Available online 7 August 2011

Keywords:
Microsatellites
Mendelian inheritance
Null alleles
Gene duplication
Inbreeding
Parasites
Cestoda
Oochoristica

ABSTRACT

Cryptic aspects of parasite population biology, e.g., mating systems, are increasingly being inferred from polymorphic and co-dominant genetic markers such as microsatellite loci. Underlying the use of such co-dominant markers is the assumption of Mendelian inheritance. The failure to meet this assumption can lead to artifactual statistics and erroneous population inferences. Here, we illustrate the importance of testing the Mendelian segregation and assortment of genetic markers and demonstrate how field-collected samples can be utilised for this purpose. To examine the reproductive mode and mating system of hermaphroditic parasites, we developed microsatellites for the cestode, *Oochoristica javaensis*. Among loci, we found a bimodal distribution of F_{IS} (a fixation index that quantifies the deviation from Hardy–Weinberg equilibrium within subpopulations) values where loci were either highly negative (close to −1) or highly positive (~0.8). By conducting tests of Mendelian segregation from natural crosses, we determined that loci with negative F_{IS} values were in fact duplicated loci that were amplified by a single primer pair. Genetic crosses also provided linkage data and indicated that the duplicated loci most likely arose via tandem duplications rather than whole genome/chromosome duplications. By correcting for the duplicated loci, we were able to correctly infer that *O. javaensis* has sexual reproduction, but the mating system is highly inbred. To assist others in testing Mendelian segregation and independent assortment from natural samples, we discuss the benefits and limitations, and provide guidelines for particular parasite systems amenable to the methods employed here.

1. Introduction

Molecular markers and population genetic analyses have greatly facilitated our ability to assess previously unobtainable data on the ecological or evolutionary dynamics of metazoan animal parasites in natural populations (Criscione et al., 2005; de Meeûs et al., 2007a). In particular, co-dominant and highly polymorphic markers such as microsatellites have provided insight into geographic structure (e.g., Louhi et al., 2010), parasite transmission patterns (e.g., Guzinski et al., 2009; Criscione et al., 2010), mating systems (e.g., Steinauer, 2009), host–race associations (McCoy et al., 2001) and hybridisation (Detwiler and Criscione, 2010). Microsatellites have also been used to understand how genetic relatedness influences life history strategy (Lagrue et al., 2009) and, in some studies, have revealed cryptic structure (Grillo et al., 2007; Criscione et al., 2011).

Underlying the use of co-dominant markers for the above-mentioned studies are the assumptions that the markers are neutral (and not linked to selected loci) and have Mendelian inheritance. Commonly, these assumptions are assessed during the development of the markers by testing loci within a population for deviations from Hardy–Weinberg equilibrium (HWE) (e.g., Criscione and Blouin, 2005). Because HWE is essentially the manifestation of Mendelian segregation at the population level, researchers often implicitly assume that markers in HWE also have Mendelian segregation. In contrast, if there are deviations from HWE, it may not be immediately apparent if non-Mendelian inheritance is the cause. This is because deviations from HWE can be caused by non-Mendelian factors such as null alleles, unrecognised duplicated loci and unrecognised sex-chromosome loci, or population level factors such as inbreeding, selection and the Wahlund effect. Due to non-Mendelian factors, de Meeûs et al. (2004) emphasised the need for "prudence" when using microsatellites to infer parasite population processes. Failure to recognise non-Mendelian factors can lead to the calculation of artifactual statistics and thus, erroneous population inferences.

One way to achieve such "prudence" is by direct tests of Mendelian segregation via the classical genetic approach of staging a cross. Indeed, genetic crosses performed by de Meeûs et al. (2004) revealed tick microsatellites with null alleles or large allele dropout and a locus that was potentially duplicated. In addition to Mendelian segregation, genetic crosses can also provide data on Mendelian (independent) assortment between loci and reveal physical linkage, which itself is critical to know for certain down-

* Corresponding author. Tel.: +1 979 845 0925; fax: +1 979 845 2891.
E-mail address: jdetwiler@mail.bio.tamu.edu (J.T. Detwiler).

0020-7519/$36.00
doi:10.1016/j.ijpara.2011.07.003

stream applications that examine mating systems (e.g., Shaw et al., 1981; Ritland, 2002). Despite the utility of genetic crosses to aid in the assessment of co-dominant markers, we are aware of only three species of metazoan parasites of animals where genetic crosses were used to statistically test Mendelian inheritance of microsatellite loci: the ticks *Ixodes ricinus* (de Meeûs et al., 2004; Røed et al., 2006) and *Bothriocroton hydrosauri* (Guzinski et al., 2008), and the blood fluke *Schistosoma mansoni* (Criscione et al., 2009). Understandably, two factors may explain the paucity of crosses. First, parental and offspring genotypes are needed. Obtaining genotypes for both parents and offspring can be invasive or impossible for many important pathogens of humans or valuable domestic animals. In addition, the small size of parasite eggs or larvae may preclude the genotyping of multiple loci per individual offspring and thus may require hundreds of offspring to test just 10–20 loci. To some degree, solutions to the latter problem have improved and newer protocols have shown that reliable multilocus genotypes can be obtained from individual parasite larvae (e.g., Sorensen et al., 2006; Steinauer et al., 2008; Valentim et al., 2009). Second, the life cycle complexity of many parasites may be an obstacle to staging crosses (e.g., maintaining hosts in the laboratory). However, this latter problem can be circumvented in some parasites, primarily endoparasites that are found at low intensities of infection, by sampling parents and their offspring from naturally infected hosts. Because adult endoparasites are present within a closed mating system, i.e., individuals cannot mate with parasites in other hosts, genetic crosses in nature can easily be recognised and exploited. As a case in point, an adult hermaphroditic fluke or tapeworm that is alone in a host must mate with itself to produce offspring.

Here, we illustrate how the ability to test for Mendelian inheritance in parasite samples collected from natural populations enabled the correct inference of the parasite's mating system. In the tapeworm *Oochoristica javaensis*, we observed highly variable F_{IS} (an index of the inbreeding of individuals resulting from the non-random union of gametes within a subpopulation) values among microsatellite loci. F_{IS} quantifies the deviation from HWE with negative values indicating an excess of heterozygotes, and positive values indicating an excess of homozygotes (see de Meeûs et al., 2007a for a thorough explanation). By conducting tests of Mendelian inheritance with offspring from natural crosses, we could test whether non-Mendelian factors or the reproductive mode/mating system were responsible for the departures from HWE. Correct inference of population level processes would not have been possible without tests of Mendelian segregation that identified duplicated loci. In addition, we were able to test for non-random assortment between loci and in one cross, provide estimates of recombination among physically linked loci. These linkage data suggest that the duplicated loci are tandem repeats and not the result of whole genome or chromosome duplications. To assist others in testing Mendelian inheritance from natural parasite samples, we discuss the benefits and limitations, and provide guidelines for particular parasite systems amenable to the methods employed here.

2. Materials and methods

2.1. Study system

Oochoristica javaensis is a cestode found in the intestines of geckos. It was first described from geckos in Indonesia (Kennedy et al., 1982), but has also been recovered from an invasive gecko species in the United States (Criscione and Font, 2001b). This hermaphroditic parasite sexually matures and mates within the intestine of the gecko. Gravid proglottids (cestode segments that contain eggs) are released into the environment with the host's faeces. An intermediate host consumes the eggs, which then develop within the body cavity into the larval stage or metacestode (precysticercus). The life cycle is perpetuated when the gecko consumes the invertebrate host, which is unknown in nature. However, the red flour beetle, *Tribolium castaneum*, is a suitable intermediate host in the laboratory (Criscione and Font, 2001a).

2.2. Microsatellite library development

A microsatellite library was developed from the genomic DNA of a single *O. javaensis* cestode using a universal linker and ligation procedure described by Hamilton et al. (1999) and modified by Grant and Bogdanowicz (2006). Enrichment of the library was with the same set of dimeric, trimeric and tetrameric oligonucleotide repeats as Barnett et al. (2008). Library construction and sequencing of clones was conducted by S. Bogdanowicz at the Evolutionary Genetics Core Facility at Cornell University, USA. A total of 184 colonies from the enriched library were sequenced, but there were only 54 unique microsatellite loci. Of these, primers were designed for 44 loci with sufficient flanking regions and desirable repeat motifs/lengths using Primer 3 (Rozen and Skaletsky, 2000).

2.3. Obtaining cestodes for genetic analysis

Mediterranean geckos, *Hemidactylus turcicus*, were collected from buildings mostly within College Station, Texas, USA, but a few came from other locations to screen for geographic variation (Supplementary Table S1). Cestodes were removed from the intestine during dissection and placed in 70% ethanol. The anterior region (neck and scolex) was used to obtain individual genotypes for population analysis of adult worms and to determine the parent genotypes in the cross data to test Mendelian inheritance. To avoid allelic contamination from potential outcrossing partners, the scolex region was used because it does not contain mature proglottids (the repeated reproductive segments of a tapeworm) that might have outcrossed sperm. Genotypes of offspring were used to investigate the Mendelian segregation of single loci and to test the association and potential linkage between loci. Offspring were collected at two different life stages. Most of the analyses were based on the metacestode stage because this stage provided ample amounts of good quality DNA for microsatellite genotyping. To obtain metacestodes, live adult cestodes with gravid proglottids were examined under a compound microscope at 200× for the presence of oncospheres with hooks. If gravid proglottids were present, eight gravid proglottids were placed in a 35 mm × 10 mm well dish with filter paper lining the bottom and lightly covered in white flour. Ten *T. castaneum* beetles were then placed in the well dish and allowed to feed on the proglottids (Criscione and Font, 2001a). After 20 days, beetles were dissected and metacestodes were removed and stored in 70% ethanol. In one cross, we also tested whether we could directly genotype offspring from the oncosphere stage (larvae inside the eggs). Forceps were used to tease out the oncospheres from the gravid proglottids after the adult worm had been stored in ethanol.

2.4. DNA extraction and microsatellite genotyping

DNA was extracted from all three tissue types with 5% chelex containing 0.2 mg/mL proteinase K. The extract volume for the scolex, metacestode and oncosphere was 200, 100 and 50 μl, respectively. Samples were incubated at 56 °C for 2 h and then boiled at 100 °C for 8 min.

We used the M13 method to genotype where a M13 oligonucleotide (TGTAAAACGACGGCCAGT) was added to the 5′-end of the forward primer (Schuelke, 2000). To reduce polyadenylation,

a 5′ sequence tag (GTTTCTT) was added to the reverse primer. PCR amplifications were performed in 15 μl reaction volumes containing 1.5 mM $MgCl_2$, 1× buffer (50 mM KCl, 10 mM Tris–Cl, 0.1% Triton X-100), 0.2 mM each of dNTP, 0.08 μm M13-forward primer, 0.16 μm reverse primer, 0.16 μm fluorescent-labelled M13 primer (Applied Biosystems: FAM, VIC, NED or PET), and 0.05 U/μl NEB-Taq polymerase (New England BioLabs). The amount of chelex-extracted genomic DNA from scoleces and metacestodes was 2.4 μl, and 5 μl for oncospheres. The thermocycler profile was 94 °C for 5 min, followed by 31 cycles of 94 °C for 30 s, 56 °C for 45 s, 65 °C for 45 s, and then nine cycles of 94 °C for 30 s, 53 °C for 45 s, 65 °C for 45 s, followed by an extension at 65 °C for 10 min. Fragments were visualised on a 3730*xl* 96-Capillary Genetic Analyzer with 500 LIZ size standard at the DNA Analysis Facility on Science Hill at Yale University, USA. Alleles were scored and manually inspected with Genotyper 3.7 (Applied Biosystems). For quality control, we regenotyped 5% of the adult dataset and independently scored the adult genotypes. We found no discrepancies.

2.5. Population analyses

The polymorphism of loci was screened in an initial subset of our samples from across at least three geographic locations (for sample sizes see Supplementary Table S1). Fourteen of these loci amplified, but showed no allelic variation in these initial screens. Thus, these 14 loci were removed from further analyses. For the remaining 30 loci, population-level patterns and measures of genetic diversity were conducted with 138 *O. javaensis* collected from 32 gecko hosts from within College Station (~15 km^2 total sampled area). Observed heterozygosity (H_O; $1 - Q_{intra}$) was calculated in Genepop 4.0 (Rousset, 2008). Gene diversity (H_S) was calculated with FSTAT 2.9.3 (Goudet, 1995) using Nei's unbiased estimator. Weir and Cockerham's estimator of F_{IS} (per locus and multilocus) was computed with Spagedi 1.3 (Hardy and Vekemans, 2002). Deviations from HWE per locus and multilocus were tested by permutating alleles among individuals 20,000 times in Spagedi (Table 1).

Table 2 lists 12 loci that show evidence of duplication, i.e., a single primer pair amplified two loci (11 of these were confirmed with cross data; see Section 3). To estimate F_{IS} on these duplicated loci as if they were single copy, individuals with more than two alleles at a locus (e.g., a heterozygote at both of the loci would appear with four alleles from a single primer pair) were counted as missing data at that locus. We recognise that this is not a "real" estimate of F_{IS}, but in practise if loci are not recognised as duplicated one might conclude samples with >2 alleles were unreliable for scoring (see Section 4). Furthermore, the analyses in Table 2 are not meant to be a reflection of any real biological process, but rather serve to illustrate the effects of unrecognised duplicated loci on F_{IS} when they are analysed as a single locus. For some of the duplicated loci, we could confirm segregation patterns in the population (via the cross data; see Section 3), and thus estimate biologically relevant F_{IS} values for one or both of the loci in the co-amplified pair (Table 3). For loci with polymorphism within College Station, we estimated a multilocus F_{IS} in Spagedi and tested for genotypic disequilibrium between pairs of loci using Genepop with Markov chain parameters of 5,000 dememorisations, 5,000 batches and 5,000 iterations. These latter tests included 15 single copy loci (Table 1) and six corrected loci from duplicated markers so that we could ascertain segregation on a per locus basis (Table 3).

2.6. Tests of Mendelian inheritance in samples collected from natural populations

In total, field-collected cestodes from six hosts (A–F; see Table 2) were used to generate the cross data. All crosses originated from College Station, except those from Host C, which came from Baton Rouge, Louisiana, USA. In all crosses, offspring genotypes were generated from metacestodes. In Host A, we also repeated tests on three loci, but using oncospheres from the single parent worm (Parent A-1.1; see Table 2 for a description of maternal parental identification codes) present in this host (Table 2). Adult worms were first genotyped at the microsatellite loci to identify informative loci for offspring genotyping (sample sizes given in Table 2). We present methods for testing Mendelian segregation and independent assortment between pairs of loci for two different scenarios from field-collected hermaphroditic endoparasites: (i) parasites

Table 1
Measures of genetic diversity for single copy loci that showed polymorphism (>1 allele) either among locations[a] or within College Station, Texas, USA. Loci for which Mendelian segregation was confirmed from cross data are shown in bold. Observed heterozygosity, H_O; expected heterozygosity, H_S; local fixation index, F_{IS}, estimated from 138 cestodes from College Station.

Locus	Total alleles across all locations[a] (n = 192)	No. of alleles in College Station (n = 138)	H_O	H_S	F_{IS}
di001	2	2	0.051	0.378	0.866[c]
di008	3	3	0.051	0.350	0.855[c]
di011	2	2	0.123	0.487	0.747[c]
di030	4	2	0.044	0.270	0.839[c]
di032	3	2	0.044	0.359	0.879[c]
di033	3	2	0.043	0.280	0.845[c]
di035	2	2	0.044	0.492	0.911[c]
di046	3	2	0.087	0.476	0.817[c]
di068	4	3	0.080	0.436	0.817[c]
di073	3	2	0.101	0.491	0.793[c]
di078	3	3	0.065	0.350	0.814[c]
di094	3	2	0.044	0.359	0.879[c]
di097	2	2	0.073	0.183	0.605[c]
di109	2	2	0.065	0.445	0.853[c]
tri001	3	2	0.036	0.129	0.719[c]
tri007[b]	2	1	0	0	NA
tri022[b]	2	1	0	0	NA

NA, not applicable.
[a] Locations were within the USA states Louisiana, Alabama and Texas.
[b] These two loci were monomorphic in College Station, Texas, but had another allele present in a different geographic location. Even though Mendelian segregation has not been tested for these two markers, we included them as single copy loci as all genotyped individuals only showed a single peak.
[c] Statistical significance at $P < 0.0001$.

Table 2
Measures of genetic diversity for duplicated loci when treated as single copy loci (see Section 3.2 for explanation). Loci for which cross data support the conclusion of duplicated loci are shown in bold. Observed heterozygosity, H_O; expected heterozygosity, H_S; local fixation index, F_{IS}, estimated from 138 cestodes from College Station, Texas, USA.

Locus	Total alleles across all locations[a] (n = 192)	No. of alleles in College Station (n = 138)	H_O	H_S	F_{IS}[b]
di005	3	1	0	0	NA
di019	6	4	1.00	0.581	−0.721[c]
di044	5	4	1.00	0.742	−0.348[c]
di047	2	2	1.00	0.500	−1.00[c]
di069	5	3	0.993	0.507	−0.957[c]
di086	4	2	1.00	0.500	−1.00[c]
di102	2	2	0.906	0.496	−0.827[c]
di131	8	6	1.00	0.602	−0.660[c]
di140	3	3	1.00	0.611	−0.636[c]
tet007	4	3	0.242	0.399	0.392[c]
tet009	2	2	1.00	0.500	−1.00[c]
tet012	4	4	0.385	0.567	0.321[c]

NA, not applicable.
[a] Locations were within the USA states Louisiana, Alabama, and Texas.
[b] In the calculation of F_{IS}, individuals with >2 alleles at a locus were considered missing data at that locus (see Section 2.4 for explanation).
[c] Statistical significance at $P < 0.0001$.

Table 3
Measures of genetic diversity for co-amplified loci that are partitioned into single loci. Segregation of alleles for these duplicated loci was determined from genetic cross data (Table 5). The two loci are distinguished with a −1 or −2 following the locus name. Observed heterozygosity, H_O; expected heterozygosity, H_S; local fixation index, F_{IS}, estimated from 138 cestodes from College Station, Texas, USA.

Locus	Total alleles across all locations[a] (n = 192)	No. of alleles in College Station (n = 138)	H_O	H_S	F_{IS}
di131-1	2	2	0.000	0.015	1[b]
di131-2	6	5	0.036	0.417	0.913[b]
di044-1	2	2	0.109	0.487	0.777[b]
di044-2	3	2	0.109	0.487	0.777[b]
di140-1	2	2	0.066	0.452	0.855[b]
tet007-1	3	2	0.043	0.381	0.886[b]

[a] Locations were within the USA states Louisiana, Alabama, and Texas.
[b] Statistical significance at $P < 0.0001$.

from single infections, i.e., one adult, and (ii) hosts with more than one adult parasite.

In the case of single worm infections, analyses are relatively straightforward. In a hermaphroditic organism that can self-fertilise, only loci that are heterozygous in the single parent are informative to test Mendelian segregation or independent assortment in the offspring of that parent. Thus, offspring were only genotyped at these parental heterozygous (or presumed heterozygous, see Section 3) loci. Three hosts (A, B and E; Table 2) had a single worm. A randomisation test of goodness-of-fit (http://udel.edu/~mcdonald/statrand.html; McDonald, 2009) was used to test the expectations for three categories of genotypes: 50% heterozygote and 25% for each homozygote. To test the hypothesis of independent assortment, a randomisation test of goodness-of-fit was used to test the nine expected genotype categories of a double-heterozygote cross (e.g., RrBb × RrBb: 6.25% RRBB, 12.5% RRBb, 6.25% RRbb, 12.5% RrBb, 25% RrBb, 12.5% Rrbb, 6.25% rrBB, 12.5% rrBb, 6.25% rrbb). This latter test assumes both markers have Mendelian segregation and independent assortment. We only performed independent assortment tests between loci if the loci had Mendelian segregation. Thus if significance was found, we could reject independent assortment (i.e., markers are likely physically linked). Significance in the above tests was assessed after 5,000 randomisations.

Single infections of hermaphroditic endoparasites also provide an excellent opportunity to generate impromptu genetic maps and estimate recombination frequencies (RFs) between markers. To illustrate this point, we conducted linkage map analysis with 11 loci using Parent A-1.1 (Tables 4 and 6). Linkage map analysis was done with JoinMap v4 software (van Ooijen, 2006, Kyzama BV, Netherlands). Data were coded according to the CP (outbreeder full-sib family) population type where the locus segregation type of the parental worm was designated hkxhk. Phase was unknown and estimated by the program. Linkage groups were generated with a recombination threshold of 30%. We used the Kosambi map-

Table 4
Cross data for single copy loci where Mendelian segregation was not rejected, $P > 0.05$. The letter refers to the host (A–F) and the numbers refer to the cestode parent out of the total number of parents within that host (e.g., 1.3 indicates the parent as one out of three total parents within that host). For comparison, parents with homozygous genotypes were reported, although Mendelian segregation could not be tested in their offspring (see Fig. 1 for explanation). Offspring were metacestodes (m) or oncospheres (o). The observed number of heterozygotes (Obs. N_{HE}) and homozygotes (Obs. N_{HO}) are provided. Two values are reported under N_{HO} if both homozygote categories (following allele order of parent genotypes) could also be tested. The P-values are determined with either exact (two categories) or randomisation (three categories) tests (see Section 2.5).

Host–cestode parent	Locus	Parent genotype	Total offspring (m or o)	Obs. N_{HE}	Obs. N_{HO}	P (2-tailed)
A-1.1	di001	120/122	69 (m)	35	13, 21	0.403
C-1.3	di008	246/256	30 (m)	15	15	1
C-2.3	di008	246/256	30 (m)	13	17	0.585
C-3.3	di008	256/256	–	–	–	–
C-1.3	di011	122/138	30 (m)	14	16	0.856
C-2.3	di011	122/122	–	–	–	–
C-3.3	di011	122/122	–	–	–	–
C-1.3	di030	233/245	30 (m)	16	14	0.856
C-2.3	di030	233/245	30 (m)	16	14	0.856
C-3.3	di030	245/245	–	–	–	–
A-1.1	di032	257/269	68 (m)	28	19, 21	0.320
A-1.1	di032	257/269	38 (o)	25	4, 9	0.0866
B-1.1	di033	113/117	95 (m)	49	24, 22	0.925
A-1.1	di035	263/265	68 (m)	30	25, 13	0.080
C-1.3	di046	356/362	30 (m)	18	12	0.362
C-2.3	di046	356/362	30 (m)	13	17	0.585
C-3.3	di046	356/356	–	–	–	–
A-1.1	di046	356/360	69 (m)	30	23, 16	0.268
A-1.1	di046	356/360	40 (o)	23	11, 6	0.379
C-1.3	di068	214/216	30 (m)	15	15	1
C-2.3	di068	214/216	30 (m)	13	16	0.711
C-3.3	di068	214/214	–	–	–	–
A-1.1	di068	196/216	69 (m)	37	11, 21	0.205
C-1.3	di073	235/251	30 (m)	19	11	0.2
C-2.3	di073	235/251	30 (m)	13	17	0.585
C-3.3	di073	251/251	–	–	–	–
A-1.1	di073	247/251	69 (m)	32	15, 22	0.429
C-1.3	di078	265/273	30 (m)	15	15	1
C-2.3	di078	265/265	–	–	–	–
C-3.3	di078	265/265	–	–	–	–
A-1.1	di094	260/264	69 (m)	29	21, 19	0.399
A-1.1	di097	161/163	69 (m)	34	19, 16	0.879
A-1.1	di097	161/163	39 (o)	16	14, 9	0.290
A-1.1	di109	202/204	69 (m)	35	14, 20	0.596
D-1.2	tri001	289/289	–	–	–	–
D-2.2	tri001	289/292	30 (m)	15	15	1

Table 5
Loci for which Mendelian segregation was rejected ($P < 0.05$). The observed ratios of offspring genotypes strongly support the conclusion of duplicated loci (e.g., no homozygotes are observed; see Section 3.2). The letter refers to the host (A–F) and the numbers refer to the cestode parent out of the total number of parents within that host. For comparison, parents with homozygous genotypes were reported, although Mendelian segregation could not be tested in their offspring (see Fig. 1 for explanation). Offspring were metacestodes (m) or oncospheres (o). The observed number of heterozygotes (Obs. N_{HE}) and homozygotes (Obs. N_{HO}) are provided. Two values are reported under N_{HO} if both homozygote categories (following allele order of parent genotypes) could also be tested. The *P*-values are determined with either exact (two categories) or randomisation (three categories) tests (see Section 2.5).

Host–cestode parent	Locus	Parent genotype	Total offspring (m)	Obs. N_{HE}	Obs. N_{HO}	*P* (2-tailed)	Locus 1 and 2 from co-amplified pair[a]
C-1.3	di005	272/276	30	30	0, 0	<0.0001	1: 272/272 2: 276/276
C-2.3	di005	272/276	30	30	0, 0	<0.0001	1: 272/272 2: 276/276
C-3.3	di005	272/276	30	30	0, 0	<0.0001	1: 272/272 2: 276/276
C-1.3	di019	195/199	30	30	0, 0	<0.0001	1: 195/195 2: 199/199
C-2.3	di019	195/199	30	30	0, 0	<0.0001	1: 195/195 2: 199/199
C-3.3	di019	195/199	30	30	0, 0	<0.0001	1: 195/195 2: 199/199
E-1.1	di019	199/207	15	15	0, 0	<0.0001	1: 199/199 2: 207/207
E-1.1	di044	137/139	15	15	0, 0	<0.0001	1:137/137 2:139/139
C-1.3	di047	172/174	30	23	7, 0	0.005	1: 172/172 2:174/Null[b]
C-2.3	di047	172/174	30	21	9, 0	0.043	1: 172/172 2: 174/Null[b]
C-3.3	di047	172/172	–	–	–	–	1: 172/172 2:Null/Null[b]
E-1.1	di069	322/334	15	15	0, 0	<0.0001	1: 322/322 2: 334/334
F-1.3	di102	265/279	30	30	0, 0	<0.0001	1: 265/265 2: 279/279
F-2.3	di102	265/265	–	–	–	–	1: 265/265 2:Null/Null
F-3.3	di102	265/279	30	30	0, 0	<0.0001	1: 265/265 2: 279/279
B-1.1	di131	213/217	95	95	0, 0	<0.0001	1:213/213 2:217/217
E-1.1	di131	213/225	15	15	0, 0	<0.0001	1: 213/213 2: 225/225
E-1.1	di140	313/321	15	15	0, 0	<0.0001	1: 313/313 2: 321/321
E-1.1	tet009	237/253	15	15	0, 0	<0.0001	1: 237/237 2: 253/253
A-1.1	tet012	214/230	69	69	0, 0	<0.0001	1: 214/214 2: 230/230

[a] This column refers to the inferred parental genotypes of the co-amplified loci under the hypothesis that these loci are duplicated.
[b] See text and Supplementary Fig. S1 for explanation.

ping function to convert RFs into map distances. The regression mapping algorithm was set to default settings (RF threshold <0.4, logarithm of odds scored (LOD) >1).

In the case where there is >1 hermaphroditic worm in a host, a mixture of progeny from self-mating and outcrossing of a parental worm can distort the expected Mendelian segregation or assortment ratios for particular genotype categories depending on the genotypes of the potential parents and mating rates between parents. However, loci that are heterozygous in the parent can still be informative for testing Mendelian segregation or assortment. Fig. 1 outlines a simple example to illustrate how such tests can be conducted. The drawback to conducting these tests from hosts with >1 worm is that not all genotype classes can always be tested (for explanation see Fig. 1). In our crosses with >1 worm, only the expected proportion of heterozygotes (0.5) or double heterozygote (0.25) against all other categories combined could be used to test Mendelian segregation or assortment, respectively. To conduct these tests we used an exact binomial test for goodness-of-fit (http://udel.edu/~mcdonald/statexactbin.html; McDonald, 2009). We had three hosts that had >1 worm (Host C with three worms, Host D with two worms, and Host F with three worms, Table 2).

3. Results

3.1. Population analyses

Of the 44 microsatellites that we tested, 14 (31.8%) were monomorphic in our initial screens. These loci may show variation with more sampling, but for the sake of efficiency we did not continue genotyping these markers. In addition, Locus di013 had five or more peaks amplify in many individuals and we could not make sense of peaks in the cross data. We did not analyse di013 as it may be a multicopy gene (>2 copies). The remaining 29 loci showed at least >1 allele across the sampling locations (Supplementary Table S1). Three of these (tri007, tri022 and di005) were monomorphic in our sample of 138 tapeworms at College Station (Table 1). Thus, we were able to estimate F_{IS} and test for deviations from HWE in 26 loci at College Station (Tables 1 and 2).

The initial striking result was that there was a bimodal distribution with nine loci showing significantly negative values (eight had $F_{IS} < -0.6$) and 17 loci with significantly positive values (15 had $F_{IS} > 0.6$) (Tables 1 and 2; Fig. 2, grey bars). The multilocus estimate of F_{IS} was significantly positive (0.088, $P < 0.0001$), which suggests

Table 6
Cross data for duplicated loci where segregation could be determined for one or both of the co-amplified loci. When parsed into single loci, Mendelian ratios ($P > 0.05$) support the conclusion of two segregating loci in a co-amplified pair. The letter refers to the host (A–F) and the numbers refer to the cestode parent out of the total number of parents within that host. For comparison, parents with homozygous genotypes were reported, although Mendelian segregation could not be tested in their offspring (see Fig. 1 for explanation). Offspring were metacestodes (m) or oncospheres (o). The observed number of heterozygotes (Obs. N_{HE}) and homozygotes (Obs. N_{HO}) are provided. Two values are reported under N_{HO} if both homozygote categories (following allele order of parent genotypes) could also be tested. The *P*-values are determined with either exact (two categories) or randomisation (three categories) tests (see Section 2.5).

Host–cestode parent	Locus	Parent genotype	Total offspring (m)	Obs. N_{HE}	Obs. N_{HO}	*P* (2-tailed)
C-1.3	di131-1	213/215	30	16	14	0.856
C-2.3	di131-1	215/215	–	–	–	–
C-3.3	di131-1	215/215	–	–	–	–
C-1.3	di131-2	219/225	30	16	14	0.856
C-2.3	di131-2	225/225	–	–	–	–
C-3.3	di131-2	225/225	–	–	–	–
D-1.2	di131-1	213/213	–	–	–	–
D-2.2	di131-1	213/213	–	–	–	–
D-1.2	di131-2	225/225	–	–	–	–
D-2.2	di131-2	221/225	30	13	17	0.585
A-1.1	di140-1	311/313	69	38	15, 16	0.719
A-1.1	di140-2	321/321	–	–	–	–
A-1.1	tet007-1	207/235	69	33	20, 16	0.763
A-1.1	tet007-2	Null/275[a]	–	–	–	–
C-1.3	di044-1	137/145	30	14	16	0.856
C-2.3	di044-1	137/137	–	–	–	–
C-3.3	di044-1	137/137	–	–	–	–
C-1.3	di044-2	139/163	30	14	16	0.856
C-2.3	di044-2	139/139	–	–	–	–
C-3.3	di044-2	139/139	–	–	–	–

[a] See Supplementary Fig. S1 for broad sense definition of a null.

a slight excess of homozygotes in the population. However, there were two interesting features that suggested something was unusual about the loci with $F_{IS} < 0.4$. First, at the loci di047, di086 and tet009, all individuals appeared to be a heterozygote for the same two alleles (for each respective locus) hence $F_{IS} = -1$ (Table 2). At the other six loci with negative F_{IS} (Table 2), all or almost all scored individuals appeared as a heterozygote, though the two alleles could be different among individuals. For example, at di044, all genotypes were either 137/139 or 145/163 ($F_{IS} = -0.348$, Table 2). Second, among the loci di019, di044, di131, di140, tet007 and tet012, there were 55 out of 828 possible genotypes (six loci × 138 individuals) that were recorded as missing data because the individual showed >2 alleles at the locus. For instance, at di044, 15 individuals had the genotype 137/139/145/163. After analyzing genetic crosses for 11 of the 12 loci in Table 2 (discussed below), we had strong evidence that the aberrant F_{IS} values (mostly negative) in Table 2 could be explained by the fact that single primer pairs were amplifying duplicated loci. Even though we do not have cross data for di086, we are confident this is also a duplicated locus given that all individuals are heterozygous for the same two alleles.

For some of the duplicated loci, we were able to parse out individual loci from the co-amplified pair via the segregation results of the cross data (Tables 3 and 6). We could then correct the genotypes in the population data set to make the duplicated loci effectively single copy loci. In doing so, the F_{IS} estimates from these corrected loci were now highly positive (range from 0.77 to 1; Table 3) and fitted within the range of single copy loci given in Table 1 (Fig. 2, black bars). The revised multilocus F_{IS} estimate (21 polymorphic loci from Tables 1 and 3) was 0.83, $P < 0.0001$, which is much greater than when duplicated loci were included and treated as single loci.

We also found significant genotypic disequilibrium among these 21 loci with 75% (158/210) of the pairwise comparisons with a $P < 0.05$. Even after a sequential Bonferroni, 120/210 remained significant. Known physically linked loci (see Section 3.2) did not weight these results as loci pairs found to have random assortment from the crosses were just as likely to show genotypic disequilibrium as loci with non-random assortment (Tables 7 and 8) (data not shown; Fisher's exact test $P = 1$).

3.2. Results of Mendelian inheritance tests

For the 15 single copy loci with polymorphism at College Station (Table 1), the hypothesis of Mendelian segregation was not rejected in the crosses (Table 4). In Parent A-1.1, Mendelian segregation at three loci (di032, di046 and di097) was not rejected with either metacestodes or oncospheres (Table 4). In contrast, Mendelian segregation was rejected for all the loci listed in Table 5. In these crosses, progeny were examined from parents that appeared heterozygous for these loci. For the loci in Table 5, except di047, all of the offspring were also heterozygotes for the same two alleles as the parent (i.e., there was no segregation between the alleles). Thus, from the data in Table 5 alone, one would conclude that either the organism has asexual reproduction or these are duplicated loci where both loci are homozygous in the parents (the latter is what we conclude, see Section 4).

The segregation patterns at di047 were an interesting case. All genotyped individuals at College Station had the genotype 172/174 (both peaks were of equal height on the electropherogram). The cross data for di047 were generated from Host C, which came from Baton Rouge. Parents C-1.3 and C-2.3 had the genotype 172/174 and Parent C-3.3 was 172/172. However, the 174 peak in the former two parents was half the height of the 172 peak. In combination with the population data, these data suggested a duplicated locus with a possible null allele at the locus presenting the 174 allele. Thus, a working hypothesis was that the genotypes of Parents C-1.3 and C-2.3 were 172/172 at di047-1 and 174/null at di047-2, whereas Parent C-3.3 was 172/172 at di047-1 and null/null at di047-2 (Table 5). Consistent with this hypothesis, we did not detect a deviation ($P > 0.05$) from the expected 50% proportion of heterozygotes in the offspring of Parents C-1.3 or C-2.3 when using the half peak height of 174 as an indicator of a 174/null heterozygote in offspring. In addition, all offspring of each of the parents had the 172 peak. Here we note that when using the term "null" with duplicated loci, we are using it in a broad sense to include the potential for homoplastic alleles between the co-amplified loci, the potential for segregation of a chromosome with the duplication and one without, or a true null where primers fail to amplify the

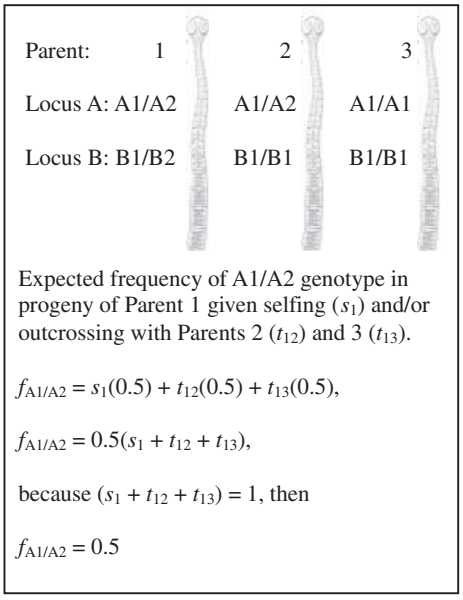

Fig. 1. An illustration of how Mendelian segregation and assortment can be tested in a cross with three worms. There are two loci (A and B) with two alleles each. In the progeny of Parent 1, the Mendelian segregation of both loci and independent assortment between the loci can be tested. Shown is a proof for the expected frequency of a heterozygote at locus A ($f_{A1/A2}$) in the progeny of Parent 1. The potential crosses expressed as proportions (which sum to 1) of sired offspring include, s_1 = the selfing rate of Parent 1 (% offspring sired by itself), t_{12} = the outcrossing rate of paternal Parent 2 to maternal Parent 1, and t_{13} = the outcrossing rate of paternal Parent 3 to maternal Parent 1. Given the genotypes in the figure, the probability of producing an A1/A2 offspring in any of these latter crosses is 0.5 (i.e., Mendelian expectation) times the respective mating rate. The sum of these probabilities equals $f_{A1/A2}$. As the Mendelian expectation is 0.5 in all of these crosses, it can be factored out. In the progeny of Parent 1, $f_{A1/A2} = 0.5$ because $(s_1 + t_{12} + t_{13} = 1)$. Thus, an exact binomial test of goodness-of-fit can be used to test for deviations from Mendelian inheritance. The drawback of this method is that not all genotype classes can be tested (see Section 4 for some exceptions). This is because the Mendelian expectation of producing an A1/A1 or A2/A2 genotype changes between the different paternal matings to Parent 1. Thus, the actual selfing and outcrossing rates would need to be known to generate expected proportions for these genotypes. Extending these arguments, the expected frequency of a double heterozygote under independent assortment will be 0.25 as any matings with Parent 1 to itself or Parents 2 or 3 have a 0.25 expectation of producing a double heterozygote. In the scenario above, only Mendelian segregation at locus A can be tested in Parent 2 and no tests can be performed with the progeny of Parent 3.

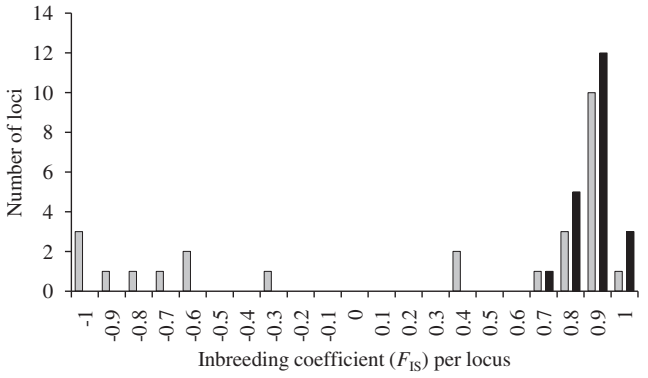

Fig. 2. The influence of duplicated loci on the distribution of population F_{IS} values. The white bars illustrate the bimodal distribution that occurred when each duplicated locus was analysed as one single copy locus (Tables 1 and 2). The black bars illustrate the unimodal and normal distribution that arises with the single copy loci (Table 1) and corrected genotypes of individual parsed loci of a co-amplified pair (Table 3; see Section 3.2 for details on correction).

allele (see Supplementary Fig. S1 for a distinction of the three possibilities).

We classified di105, also tested in Host C of Baton Rouge, as a duplicated locus as all three parents and their offspring had the peaks 272 and 276 (Table 5). The interesting aspect of this locus was that it was monomorphic for a 268 allele at College Station. We included this locus in Table 2, but our current data do not allow us to distinguish whether this locus in College Station is monomorphic single copy, duplicated and homozygous for the same allele at both loci or duplicated with null homozygotes at one of the loci.

We obtained cross data for parents with >2 alleles at a locus (Table 6). The following five sets of cross data allowed us to ascertain allelic segregation at individual loci within four co-amplified pairs: (1) di131 in Parent C-1.3 had the genotype 213/215/219/225 and the other two parents had 215/225, (2) di131 in Parent D-2.2 had the genotype 213/221/225 and Parent D-1.2 had 213/225, (3) di140 in Parent A-1.1 had 311/313/321, (4) tet007 in Parent A-1.1 had 207/235/275, and (5) di044 in Parent C-1.3 had 137/139/145/163 and the other two parents were 137/139. In Table 6, a non-significant *P*-value indicated that alleles segregated in Mendelian fashion for each of the loci in the co-amplified pairs. For example, with di131 in Parent C-1.3, alleles 213 and 215 (locus di131-1) and 219 and 225 (locus di131-2) segregated in Mendelian fashion. Interestingly, with tet007, 235 and 275 were always inherited together but had Mendelian segregation with allele 207 in Parent A-1.1. Thus, we designated 207/235 as locus tet007-1 and "null"/275 as tet007-2 (see Supplementary Fig. S1).

Independent assortment between loci was tested with offspring from four different parents (A-1.1, C-1.3, C-2.3 and D-2.2). A total of 11, 11, five and two loci were heterozygous within Parents A-1.1, C-1.3, C-2.3 and D-2.2, respectively. All pairwise combinations between loci within a cross were tested. We included duplicated loci in these analyses if we could determine allele segregation as presented above. For example, in Parent C-1.3, we included di044-1, di044-2, di131-1 and di131-2. The bi-locus genotypic proportions were significantly different from Mendelian expectations in 14/121 pairwise comparisons, thus rejecting independent assortment among some pairs of loci (Tables 7 and 8, non-significant results not shown). Between one pair of loci, di046 and di073, data from three crosses (Parents A-1.1, C-1.3 and C-2.3) showed non-random assortment.

The linkage map analysis from Parent A-1.1 assembled the loci into three linkage groups that consisted of two, three and four loci (Fig. 3). Loci di068 and di097 did not have any linkage to other loci. The results of the physical map corresponded to the independent assortment tests in that pairs of loci with significant non-random assortment were estimated to be in the same linkage group (compare Fig. 3 with Table 7). The one exception was with tet007-1 (see Fig. 3 legend).

4. Discussion

4.1. Teasing apart the presence of duplicated loci from alternative modes of reproduction/mating systems

We observed an unexpected, but intriguing pattern of bimodal F_{IS} values among microsatellite loci in a population of *O. javaensis* (Fig. 2, grey bars). Because these markers were generated randomly from an enriched library and because there are several loci with either highly positive and or highly negative F_{IS} values, it did not seem plausible to invoke selection or assortative/disassortative mating as these mechanisms only affect loci (or linked loci) associated with the trait(s) under selection/mate choice. It is also unlikely that a given mating system (random mating versus inbreeding) alone would create such a large variance in F_{IS} among loci. Thus, can different modes of reproduction (asexual versus sexual) explain this pattern among assumed neutral loci?

Table 7
Pairwise comparisons between loci with non-random assortment from Parent A-1.1. The nine categories of observed bi-locus genotype proportions were significantly different from the proportions expected for two independently assorting loci ($P < 0.05$).

Host–cestode parent	Loci	Obs RRBB	Obs RRBb	Obs RRbb	Obs RrBB	Obs RrBb	Obs Rrbb	Obs rrBB	Obs rrBb	Obs rrbb	*P* value (2-tailed)
A-1.1	di032 vs. di094	0	0	19	0	28	0	21	0	0	<0.0001
A-1.1	di032 vs. di109	3	13	3	7	17	4	4	5	12	0.006
A-1.1	di094 vs. di109	4	5	12	7	17	5	3	13	3	0.0104
A-1.1	di001 vs. di140	0	5	8	3	25	7	12	8	1	<0.0001
A-1.1	di035 vs. di046	17	7	1	5	19	6	1	3	9	<0.0001
A-1.1	di035 vs. di073	1	6	18	4	23	3	10	2	1	<0.0001
A-1.1	di046 vs. di073	0	2	21	1	28	1	14	2	0	<0.0001
A-1.1	di109 vs. tet007	0	2	12	3	28	4	17	3	0	<0.0001

vs, versus.

Table 8
Pairs of loci from Parents C-1.3 and C-2.3 for only a single category of genotypic proportions, the observed double heterozygotes (N_{HE}). Not all nine categories could be tested due to the presence of >1 worm in a host (see Fig. 1).

Host–cestode parent	Loci	Obs. double N_{HE}	Obs. other categories combined	*P* value (2-tailed)
C-1.3	di011 vs. di044-1	14	16	0.010
C-1.3	di011 vs. di044-2	14	16	0.010
C-1.3	di044-1 vs. di044-2	16	14	0.001
C-1.3	di046 vs. di073	18	12	<0.0001
C-1.3	di131-1 vs. di131-2	16	14	0.001
C-2.3	di046 vs. di073	13	17	0.032

vs, versus.

Taken at face value, the large variation in F_{IS} among loci (S.E. = 0.15; Fig. 2, grey bars) together with the fact that we found many identical multilocus genotypes among individual cestodes (unpublished data) closely matches the theoretical predictions of a 0.99 to 0.9999 rate of clonal reproduction (Balloux et al., 2003; de Meeûs et al., 2006). Also, at first glance, all of the crosses (except di047) in Table 5 suggest asexual reproduction from parent to offspring. However, a strongly negative F_{IS} value is expected with high asexual reproduction (Balloux et al., 2003). In contrast to this expectation, we observed a mean F_{IS} of 0.22 among loci in Fig. 2 (grey bars) and the multilocus estimate was 0.088. Furthermore, we observed that within some crosses (in Hosts A, B, and C) there could be loci with Mendelian segregation and other loci with apparent asexual inheritance (Tables 4 and 5). For example, in Parent B-1.1, di033 had Mendelian segregation but all of the offspring had the same maternal heterozygote genotype at di131. These latter crosses rule out complete asexual reproduction from parent to offspring. However, under some forms of automictic parthenogenesis (unfertilised eggs are diploid, although meiosis occurs during oocyte formation), it is possible to have some loci segregating and other loci with asexual inheritance (see Fig. 2 in de Meeûs et al., 2007b). Nevertheless, we feel that automictic parthenogenesis is unlikely for the following reason. Several loci (di131, di140, di044) that showed presumed asexual inheritance with just two alleles in the parent of some crosses were shown to have Mendelian

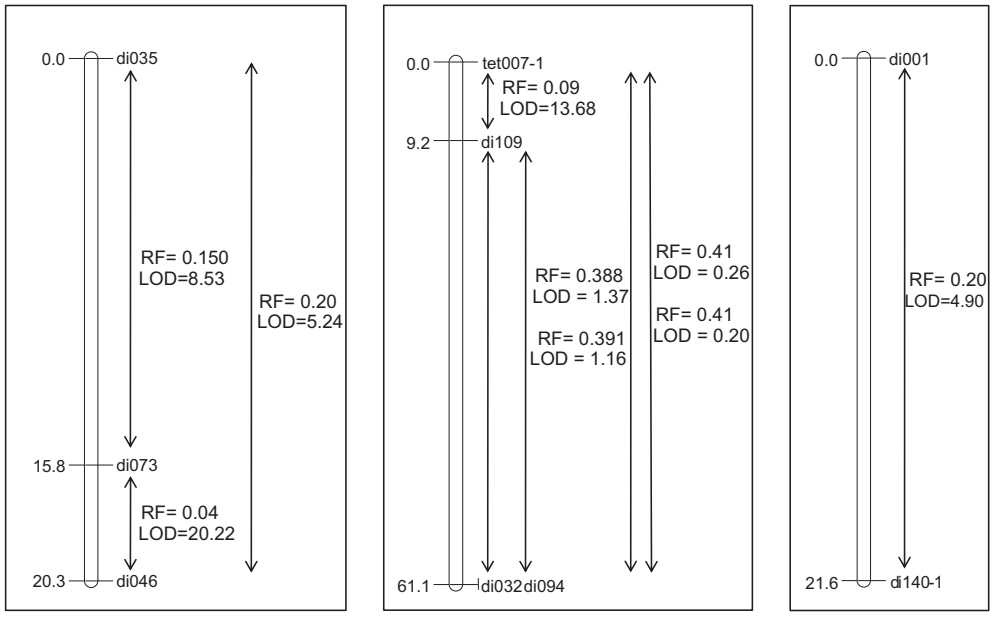

Fig. 3. Three linkage groups identified from the cross of Parent A-1.1. Loci di068 and di097 did not have any linkage to other loci and thus are not shown. The numbers on the left indicate the cumulative map distance in centimorgans. Maximum likelihood estimates of pairwise recombination frequencies (RF) with their respective logarithm of odds (LOD) scores are shown with arrows between loci. No recombination was detected between di032 anddi094. We had to reduce stringency in the regression mapping algorithm of JoinMap v4 in order to join tet007-1 and di109 into the same linkage group with di032 and di094. We did this largely for illustrative purposes as di109, which was linked to tet007-1, did show non-random assortment to di032 and di094 whereas tet007-1 did not (Table 7). The RF between di109 and either di032 or di094 was 39%, but the LOD score (log-likelihood ratio comparing the estimated value of the pairwise RF with 0.5) was <1.5. A LOD of 3 is often considered a minimum for significance, so we cannot reject free recombination between di109 and di032 or di094. Thus, the assignment of tet007-1 and di109 to the same linkage group as di032 and di094 is tentative (Fig. 3).

segregation when >2 alleles were present in the parents of other crosses (Tables 5 and 6). We never observed clonal-like inheritance when more than two alleles were present in the parent (i.e., all alleles in the parent would be found in each offspring) as there were always some alleles segregating (including di047 in Host C when treating the null as a third allele). These results strongly suggest that the apparent asexual inheritance is really an artifact of the parent being homozygous at the two loci in a co-amplified pair.

Further support for duplicated loci comes from the fact that we could correct genotypes in the population sample using segregation patterns from the cross data (Table 5). Once corrected, each of the co-amplified loci had F_{IS} values that fell within the range of the single copy loci (Tables 1 and 2). Indeed, among the 21 polymorphic loci at College Station (Tables 1 and 3), a normal distribution of F_{IS} values (mean = 0.83, S.E. = 0.018) emerges (Fig. 2, black bars). The large positive multilocus estimate of F_{IS} is evidence that *O. javaensis* has a highly inbred mating system. The high amount of genotypic disequilibrium that was observed among loci is also consistent with an inbred species as the linkage disequilibrium rate of decay is reduced in inbred mating systems (Hedrick, 2005). Future studies will explore what life history traits of *O. javaensis* drive this high level of inbreeding.

4.2. Recognising duplicated loci from a population sample

It is evident from our data that unrecognised duplicated loci tend to have negative F_{IS} values. The reason is that individuals appear to be heterozygotes when at least one allele is amplifying from each of the co-amplified loci. However, the magnitude of the F_{IS} values will vary depending on the polymorphism at the co-amplified loci. In an inbred species such as *O. javaensis*, each locus of a co-amplified pair has a high chance of being homozygous, which leads to the appearance of fixed heterozygosity. For example, at di047, di086 and tet009, the same two peaks were present in each individual, thus leading to a F_{IS} value of −1. The loci di102 (F_{IS} = −0.827) and di069 (F_{IS} = −0.957) were similar in that only two peaks were observed, but a few individuals had only one peak, which may indicate a null allele at one of the co-amplified loci or a segregating duplication (Supplementary Fig. S1). In contrast, all individuals were apparently heterozygous, but not necessarily for the same two peaks at loci di019, di044, di131 and di140. A range of negative F_{IS} values was observed among these more polymorphic loci with the least negative at di044 (−0.348), where all scored individuals were either 137/139 or 145/163. We refer readers to Balloux et al. (2003) for a mathematical treatment of how excess heterozygosity and locus polymorphism influences negative F_{IS} values (see also de Meeûs et al., 2006; Simo et al., 2010). Furthermore, a null allele at one of the co-amplified loci or segregation of the duplication itself can lead to positive F_{IS} values. Indeed, this scenario was confirmed with cross data (Table 6) for locus tet007 (F_{IS} = 0.392). Examination of the population data suggests a similar mode of inheritance for tet012 (F_{IS} = 0.321), but we need cross data for this locus before attempting to parse out segregating alleles of the co-amplified loci. Thus, a highly negative F_{IS} value is indicative of an unrecognised duplicated locus, but not necessarily a prerequisite.

Another indication of duplicated loci was the occurrence of individuals (46/138; 33%) that showed more than two alleles at one or more of the following loci: di019, di044, di131, di140, tet007 and tet012. Due to the low polymorphism (likely due to inbreeding) among the loci of *O. javaensis*, however, the frequency of genotypes with more than two alleles was low (55/828; 6.6%) among these six loci. If we had not genotyped a large number of individuals, these allele patterns may not have been readily observed. Anecdotally, it is not unheard of to regard such genotypes as missing data as a locus genotype with more than two alleles in an individual might be considered contamination (PCR or template) and/or genotyping error. Thus, in a low polymorphic species, loci presenting more than two alleles in an individual should be subject to additional investigation such as genetic crosses if possible. Caution is also advised in a highly polymorphic species where the opposite pattern of many individuals at many loci showing more than two alleles could be confused with ploidy levels greater than 2*n*.

The cumulative data based on the distribution of F_{IS} values (especially with highly negative values) and the presence of loci displaying genotypes with more than two alleles in an individual suggest the potential presence of duplicated loci. However, it is not known how efficient population level analyses alone are in identifying duplicated loci in metazoan parasites of animals as the use of microsatellites is still in its infancy among these organisms. We do call attention to recent studies on the trematode *Coitocaecum parvum*, which is known to self-mate due to gravid individuals still encysted in their intermediate hosts (Lagrue et al., 2007, 2009). In this species, 12 microsatellites had F_{IS} values from 0.726 to 0.985, whereas two loci had −0.931 and −0.764 (Cpa-3 and Cpa-4, respectively), with over 95% of the individuals observed to be heterozygotes at these two loci (Lagrue et al., 2009). The authors suggested associative overdominance (an increase in fitness of heterozygotes at a neutral locus because it is in gametic disequilibrium at a locus that is under selection) as a potential cause of negative F_{IS} values and proceeded to include these two loci in subsequent analyses. Given the patterns discussed above, we feel a more parsimonious conclusion is that these loci are really unrecognised duplicated loci where the co-amplified pairs are largely homozygous due to a large degree of inbreeding (as suggested by the other 12 loci and biology of the organism). The incorporation of unrecognised duplicated loci can have serious consequences for data analyses and subsequent interpretation of results. For example, if these two loci reported by Lagrue et al. (2009) are duplicated, then the current multilocus F_{IS} value is biased downward and the calculations for the probability of repeated clonal genotypes via sexual reproduction are based on artifactual loci.

4.3. Genotyping oncospheres

Genotyping oncospheres from tapeworms could present two problems: (i) maternal vitellocytes, which contain DNA, are incorporated into the outer embryonic envelope of the oncosphere (Conn and Świderski, 2008). Thus, it is possible that maternal DNA is present in the oncospheres. If this were the case, then all offspring would have the same heterozygote genotype as the maternal parent. All three loci tested in Parent A-1.1 showed Mendelian segregation (Table 4), thus it does not appear maternal DNA persists in the oncosphere of *O. javaensis*. (ii) The oncospheres are small (~20–30 μm in length and width; Criscione and Font, 2001c) and thus limited amounts of DNA are available for downstream PCR. Such limited amounts could lead to allelic dropout and, thus, the appearance of excess homozygotes. Again, the three tested loci in Parent A-1.1 had Mendelian segregation (Table 4), so it does not appear that this was a problem. However, we note that we only had approximately 42% success in obtaining genotypes from oncospheres. We extracted DNA from 96 oncospheres but over half of these did not amplify for any loci. Thus, individual oncospheres either worked or did not work. We are currently investigating whether whole genome amplification methods (e.g., Valentim et al., 2009) might improve amplification success.

4.4. Tests of independent assortment and the origin of the duplications

Methods that examine hermaphroditic mating systems or parentage in general often work under the assumption that loci are

independent in order to estimate outcrossing rates or probabilities of parentage assignment across loci (Shaw et al., 1981; Ritland, 2002; Jones et al., 2010). Thus, knowledge of physical linkage among loci becomes imperative. For example, one would not want to use di032 and di094 (Fig. 3) simultaneously in the same progeny array to estimate selfing rates. However, the joint use of di032 and di109 would be appropriate given their map distance exceeds 50 cM.

Here we illustrated the ability to test non-random assortment from field-collected parasites. In doing so, we can now appropriately choose markers for future studies to examine the primary selfing rates in *O. javaensis*. Moreover, these tests of non-random assortment elucidate a likely origin of the duplicated loci. If the loci are the result of genome-wide or whole chromosome duplication events, then we would expect duplicated loci to assort independently. However, in Host C, the co-amplified loci at di044 showed significant non-random assortment as did the co-amplified loci at di131 (Table 8). The co-amplified loci of tet007 also show evidence of linkage in that allele 235 is always present with 275 in the cross data (Table 6) and in the population data (not shown). Thus, this physical linkage suggests these loci originated as tandem duplications. Cursory, but not definitive, support for a lack of whole chromosome/genome duplication is given by the result that single copy loci showed linkage to duplicated loci (e.g., di009 and tet007-1; di001 and di140-1; di011 to both di044-1 and di044-2; Table 3, Fig. 3). This latter result also suggests duplications are not confined to one particular part of the genome.

The preponderance of duplicated microsatellite loci suggests several rarely explored research areas to pursue with inbred flatworm parasites. A high percentage of screened loci were duplicated (12 of the 44, 27.3%), which contrasts with the hypothesis by Wright et al. (2008) that inbred organisms should have reduced genome sizes. Further, the timing of the duplications is unknown, although they may have recently arisen since most of the duplicated loci had low repeat differences (David et al., 2003). Perhaps most importantly, duplications have led to adaptive novelty in parasites (Nair et al., 2007), thus it will be interesting to see if other inbred flatworms show similar patterns of gene duplication.

4.5. Guidelines for testing Mendelian inheritance from field-collected parasites

Our goal is not to suggest that all co-dominant markers for every species be tested for Mendelian inheritance as we recognise such tests will not be feasible for all metazoan parasites of animals. Rather we hope these guidelines will aid researchers in recognising applicable systems, especially when deviations from HWE are detected in a population sample. These guidelines are not exhaustive, but rather provide a simplified framework for those not familiar with genetic crosses.

In general, tests of Mendelian inheritance require genotypes of known parents and their offspring (approximately 30 or more). Because parents can be more readily recognised in closed mating systems, the methods we present will most likely be applicable to endoparasites (where the host defines a discrete mating boundary) compared with ectoparasites. However, this will depend on the specific biology of the ectoparasite in question. A hermaphroditic parasite (e.g., most parasitic platyhelminths) that is alone in a host represents an ideal situation as the maternal and paternal genotypes are determined from that single parasite. Thus, all heterozygous loci can be tested for Mendelian segregation and assortment in all genotype categories and linkage maps are easily created. An analogous situation is present in a dioecious species where there is a single male and any number of females as long one can obtain parasite offspring directly from the maternal parents. Only one individual of the dioecious pair needs to be a heterozygote at a locus or a double heterozygote at two loci to test segregation and assortment, respectively.

When two or more paternal parasites are present (e.g., two hermaphroditic parasites in a host), it may not be possible to test all genotype categories. Obviously this reduces power to detect deviations from Mendelian segregation or independent assortment. Nevertheless, it is possible to still test some expectations based on Mendelian inheritance, depending on the genotypes of the potential parents and without knowing complete parentage (see example in Fig. 1). The example in Fig. 1 is also applicable to dioecious species. For example, the same 0.5 expectation for the frequency of an A1/A2 genotype holds in the offspring of Parent 1 if it is female and the other two parents are males. Even if more than one paternal parent is present in a host, it may still be possible to test more than two genotype categories. For instance, if all three worms in Fig. 1 at locus A were A1/A2, then three genotype categories can be tested in the offspring of any of these parents. Furthermore, imagine a third locus where each parent did not share any alleles. Such a locus could allow complete exclusion of parents (see Jones et al., 2010), and thus the reconstructed full sibling relationships could be used to test genotype categories at other loci. However, this exclusion method should not be confused with probabilistic methods that reconstruct sibling relationships. The latter would be an inappropriate approach as probabilistic methods assume Mendelian inheritance (Jones et al., 2010) and thus do not provide an independent means to test segregation patterns.

As more parental parasites are added to a host, testing Mendelian inheritance as given in Fig. 1 will likely become intractable (mainly due to the need to genotype many potential parents). However, many endoparasitic metazoan parasites have aggregated distributions (Shaw et al., 1998) and, thus, finding low intensities of infection of three or four parasites are not uncommon. Allelic polymorphism should not be a limitation as we have shown here that such tests are feasible even in a highly inbred organism. However, more polymorphism will generally make things easier. Lastly, one caveat may be that a paternal parent is no longer present in the host at the time of sampling. This might be revealed if a non-maternal allele that did not match any other paternal parent was found in the offspring or if all loci show deviations from Mendelian expectations. Additional crosses should be examined in such a case or potentially the methods in Fig. 1 could be employed to test fewer genotype categories. This latter approach may also make some ectoparasite systems more tractable in the absence of being able to sample all potential paternal parents.

4.6. Summary

In studying the population genetics of the tapeworm *O. javaensis*, we observed a bimodal distribution in the population F_{IS} values among 26 microsatellite loci. Our approach illustrates the pitfalls associated with presenting and interpreting average F_{IS} values over multiple loci and demonstrates the importance of examining the results for individual loci. By taking advantage of the fact that hosts represent closed mating systems for endoparasites, we were able to exploit natural crosses to test Mendelian inheritance. These tests were necessary to tease apart the presence of duplicated loci from plausible reproductive modes that could also cause the bimodal distribution. By correcting for the duplicated loci, we were able to correctly infer that *O. javaensis* has a sexual reproductive mode but the mating system is highly inbred. Genetic cross data also provided linkage data and indicated that the duplicated loci most likely arose via tandem duplications rather than whole genome/chromosome duplications. We discuss what genotype patterns (e.g., fixed heterozygosity, individuals with more than two alleles at a locus) give clues to detecting duplicated loci from population samples, although cross data are recommended to con-

firm inheritance. Lastly, we discuss guidelines to assist others in testing Mendelian inheritance of co-dominant markers from field-collected parasite crosses.

Acknowledgements

We thank Johnson Cheng, Emily Kasl and Shih-Jye Yei for assistance with field collections. Steve Bogdanowicz at the Evolutionary Genetics Core Facility at Cornell University, USA developed the microsatellite library.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpara.2011.07.003.

References

Balloux, F., Lehmann, L., de Meeûs, T., 2003. The population genetics of clonal and partially clonal diploids. Genetics 164, 1635–1644.

Barnett, J.R., Stenzler, L.M., Ruiz-Gutierrez, V., Bogdanowicz, S.M., Lovette, I.J., 2008. Isolation and characterization of microsatellite markers from the white-ruffed manakin *Corapipo altera* (Aves, Pipridae). Mol. Ecol. Resour. 8, 215–218.

Conn, D.B., Świderski, Z., 2008. A standardised terminology of the embryonic envelopes and associated developmental stages of tapeworms (Platyhelminthes: Cestoda). Folia Parasit. 55, 42–52.

Criscione, C.D., Anderson, J.D., Sudimack, D., Subedi, J., Upadhayay, R.P., Jha, B., Williams, K.D., Williams-Blangero, S., Anderson, T.J.C., 2010. Landscape genetics reveals focal transmission of a human macroparasite. PLOS Neglect. Trop. Dis. 4, e665.

Criscione, C.D., Blouin, M.S., 2005. Eleven polymorphic microsatellite loci for the salmonid trematode *Plagioporus shawi*. Mol. Ecol. Notes 5, 562–564.

Criscione, C.D., Font, W.F., 2001a. Development and specificity of *Oochoristica javaensis* (Eucestoda: Cyclophyllidea: Anoplocephalidae: Linstowiinae). Com. Parasitol. 68, 149–155.

Criscione, C.D., Font, W.F., 2001b. The guest playing host: colonization of the introduced Mediterranean gecko, *Hemidactylus turcicus*, by helminth parasites in southeastern Louisiana. J. Parasitol. 87, 1273–1278.

Criscione, C.D., Font, W.F., 2001c. Artifactual and natural variation of *Oochoristica javaensis*: statistical evaluation of in situ fixation. Comp. Parasitol. 68, 156–163.

Criscione, C.D., Poulin, R., Blouin, M.S., 2005. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. Mol. Ecol. 14, 2247–2257.

Criscione, C.D., Valentim, C.L.L., Hirai, H., LoVerde, P.T., Anderson, T.J.C., 2009. Genomic linkage map of the human blood fluke *Schistosoma mansoni*. Genome Biol. 10, R71.

Criscione, C.D., Vilas, R., Paniagua, E., Blouin, M.S., 2011. More than meets the eye: detecting cryptic microgeographic population structure in a parasite with a complex life cycle. Mol. Ecol. 20, 2510–2524.

David, L., Blum, S., Feldman, M.W., Lavi, U., Hillel, J., 2003. Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. Mol. Biol. Evol. 20, 1425–1434.

Detwiler, J.T., Criscione, C.D., 2010. An infectious topic in reticulate evolution: introgression and hybridization in animal parasites. Genes 1, 102–123.

de Meeûs, T., Humair, P., Grunau, C., Delaye, C., Renaud, F., 2004. Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. Int. J. Parasitol. 34, 943–950.

de Meeûs, T., Lehmann, L., Balloux, F., 2006. Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. Infect. Genet. Evol. 6, 163–170.

de Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Bousses, S., Renaud, F., 2007a. Population genetics and molecular epidemiology or how to "debusquer la bete". Infect. Genet. Evol. 7, 308–332.

de Meeûs, T., Prugnolle, F., Agnew, P., 2007b. Asexual reproduction: genetics and evolutionary aspects. Cell. Mol. Life Sci. 64, 1355–1372.

Grant, J.B., Bogdanowicz, S.M., 2006. Isolation and characterization of microsatellite markers from the panic moth, *Saucrobotys futilalis* L. (Lepidoptera: Pyralidae: Pyraustinae). Mol. Ecol. 106, 253–260.

Grillo, V., Jackson, F., Cabaret, J., Gilleard, J.S., 2007. Population genetic analysis of the ovine parasitic nematode *Teladorsagia circumcincta* and evidence for a cryptic species. Int. J. Parasitol. 37, 435–447.

Goudet, J., 1995. FSTAT (Version 1.2): a computer program to calculate F-statistics. J. Hered. 86, 485–486.

Guzinski, J., Bull, C.M., Donnellan, S.C., Gardner, M.G., 2009. Molecular genetic data provide support for a model of transmission dynamics in an Australian reptile tick, *Bothriocroton hydrosauri*. Mol. Ecol. 18, 227–234.

Guzinski, J., Saint, K.M., Gardner, M.G., Donnellan, S.C., Bull, C.M., 2008. Development of microsatellite markers and analysis of their inheritance in the Australian reptile tick, *Bothriocroton hydrosauri*. Mol. Ecol. Resour. 8, 443–445.

Hamilton, M.B., Pincus, E.L., DeFiore, A., Fleischer, R.C., 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques 27, 500–507.

Hardy, O.J., Vekemans, V., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2, 618–620.

Hedrick, P.W., 2005. Genetics of Populations. Jones and Bartlett Publishers, Boston and Toronto and London and Singapore.

Jones, A.G., Small, C.M., Paczolt, K.A., Ratterman, N.L., 2010. A practical guide to methods of parentage analysis. Mol. Ecol. Resour. 10, 6–30.

Kennedy, M.J., Killick, L.M., Beverley-Burton, M., 1982. *Oochoristica javaensis* n. sp. (Eucestoda: Linstowiidae) from *Gehyra mutilata* and other gekkonid lizards (Lacertilia: Gekkonidae) from Java, Indonesia. Can. J. Zool. 60, 2459–2463.

Lagrue, C., McEwan, J., Poulin, R., Keeney, D.B., 2007. Co-occurrences of parasite clones and altered host phenotype in a snail–trematode system. Int. J. Parasitol. 37, 1459–1467.

Lagrue, C., Poulin, R., Keeney, D.B., 2009. Effects of clonality in multiple infections on the life-history strategy of the trematode *Coitocaecum parvum* in its amphipod intermediate host. Evolution 63, 1417–1426.

Louhi, K.-R., Karvonen, A., Rellstab, C., Jokela, J., 2010. Is the population genetic structure of complex life cycle parasites determined by the geographic range of the most motile host? Infect. Genet. Evol. 10, 1271–1277.

McCoy, K.D., Boulinier, T., Tirard, C., Michalakis, Y., 2001. Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. J. Evol. Biol. 14, 395–405.

McDonald, J.H., 2009. Handbook of Biological Statistics, second ed. Sparky House Publishing, Baltimore, Maryland.

Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.-C., Krishna, S., Nosten, F., Anderson, T.J.C., 2007. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. Mol. Biol. Evol. 24, 562–573.

Ritland, K., 2002. Extensions of models for the estimation of mating systems using n independent loci. Heredity 88, 221–228.

Røed, K.H., Hasle, G., Midthjell, V., Skretting, G., Leinaas, H.P., 2006. Identification and characterization of 17 microsatellite primers for the tick, *Ixodes ricinus*, using enriched genomic libraries. Mol. Ecol. Notes 6, 1165–1167.

Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Rozen, S., Skaletsky, H.J., 2000. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, New Jersey.

Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18, 233–234.

Shaw, D.J., Grenfell, B.T., Dobson, A.P., 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology 117, 597–610.

Shaw, D.V., Kahler, A.L., Allard, R.W., 1981. A multilocus estimator of mating system parameters in plant populations. Proc. Natl. Acad. Sci. USA 78, 1298–1302.

Simo, G., Njiokou, F., Tume, C., Lueong, S., de Meeûs, T., Cuny, G., Asonganyi, T., 2010. Population genetic structure of Central African *Trypanosoma brucei gambiense* isolates using microsatellite DNA markers. Infect. Genet. Evol. 10, 68–76.

Sorensen, R.E., Rodrigues, N.B., Oliveira, G., Romanha, A.J., Minchella, D.J., 2006. Genetic filtering and optimal sampling of *Schistosoma mansoni* populations. Parasitology 133, 443–451.

Steinauer, M.L., 2009. The sex lives of parasites: investigating the mating system and mechanisms of sexual selection of the human pathogen *Schistosoma mansoni*. Int. J. Parasitol. 39, 1157–1163.

Steinauer, M.L., Agola, L.E., Mwangi, I.N., Mkoji, G.A., Loker, E.S., 2008. Molecular epidemiology of *Schistosoma mansoni*: a robust, high-throughput method to assess multiple microsatellite markers from individual miracidia. Infect. Genet. Evol. 8, 68–73.

Valentim, C.L.L., LoVerde, P.T., Anderson, T.J.C., Criscione, C.D., 2009. Efficient genotyping of *Schistosoma mansoni* miracidia following whole genome amplification. Mol. Biochem. Parasit. 166, 81–84.

Wright, S.I., Ness, R.W., Foxe, J.P., Barrett, S.C.H., 2008. Genomic consequences of outcrossing and selfing in plants. Int. J. Plant Sci. 169, 105–118.